Sec. 1

Vol. 47 DECEMBER, 1956 No. 8

The Journal of Educational Psychology

3884
5 - APR 24
Copy ✓ 1957

CONTENTS

An Evaluation of a Student-Centered and Instructor-Centered Method of Conducting a Graduate Course in Education 449
GLEN R. RASMUSSEN

Theory Development and the Study of Teacher Behavior 462
DAVID G. RYANS

Class and Sex Differences in Verbal Facility of Very Bright Children. 476
HELEN H. DAVIDSON AND DOM BALDUCCI

An Experimental Comparison of Nonsense, Word, and Sentence Materials in Early Typing Training 481
LEONARD J. WEST

Behavioral Control and Intellectual Achievement of Secondary School Boys 490
NORMAN TALLENT

Book Reviews 504

Published monthly except June to September
$7.00 a year in U. S. and Pan America; Canada, $7.20; other countries, $7.40.
Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879.

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

Wm. Clark Trow, *Chairman*
University of Michigan

Stephen M. Corey
Teachers College
Columbia University

J. B. Stroud
State University of Iowa

*Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

Paul Blommers
Chester W. Harris
Harold E. Jones
H. H. Remmers
Percival M. Symonds
Miles A. Tinker
Alexander G. Wesman
Paul A. Witty

The Journal of Educational Psychology is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The Journal will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

Information for Contributors

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Wm. Clark Trow, School of Education, University of Michigan, Ann Arbor, Mich.

The Journal has set regulations regarding content and style of material published, and these should be observed in the preparation of manuscripts to be submitted.

Tables and graphs.—In keeping with the practice of most professional periodicals of like character, The Journal now requires authors to bear part of the added cost resulting from the use of tables, formulas, and graphs. Prospective contributors may obtain detailed statement as to

(Continued on Inside Back Cover)

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 47 December, 1956 Number 8

AN EVALUATION OF A STUDENT-CENTERED AND INSTRUCTOR-CENTERED METHOD OF CONDUCTING A GRADUATE COURSE IN EDUCATION

GLEN R. RASMUSSEN

School of Education, University of Michigan

This study was designed to investigate the effectiveness of two learning situations in achieving the following goals: (1) to help students learn as much as possible about present knowledge in the area of child psychology, and (2) to help students use this knowledge to guide their own behavior. In this situation the mere acquisition of data and generalizations seemed to be the less important of the two goals to the instructor. It is, of course, much easier to evaluate the success of a course from the standpoint of the first goal, acquisition, than from the standpoint of the second goal, the ability and willingness to apply classroom learning to one's everyday performance in dealing with children.

In the light of the instructor's goals for the course, what evidence can be found concerning the relative effectiveness of different teaching methods? Research attempting to measure "acquisition of subject matter" has shown conflicting results. Most of this research has been conducted in general psychology classes. Two major obstacles have been apparent. First, students in the student-centered classes were *told* they were to be student-centered. They had no choice in the situation. This is a rather autocratic and inauspicious beginning for a democratic experience. Second, how do we measure the kind of learning induced by such dissimilar methods with a paper and pencil test? A test designed by the instructor of a student-centered class violates the philosophy upon which the entire class is based. The instructor is once more the "keeper of the knowledge." Some experimenters (*1*) have felt

this conflict and informed the student-centered groups that this "final exam" would not count toward a grade. If this examination has grade meaning for one group and does not have grade meaning for the other, the factor of differential motivation is introduced and the results are subject to doubt.

The second goal of the instructor (to help students use course knowledge to guide their own behavior) is even more difficult to measure. Research in this area is practically nonexistent. There is considerable evidence that student-centered psychology classes result in significantly greater emotional and social adjustmental gains for students than do instructor-centered classes (1–3, 5–7). In this study no attempt was made to measure adjustmental gains. Student comments both during and after class suggested that the student-centered class was definitely superior in facilitating social adjustment. It would seem logical to assume that if greater emotional and social adjustmental gains were made by menbers of the student-centered sections in this study, such gains would be accompanied by an increased ability to use modern psychological principles in their own teaching.

METHOD

The data for this study were gathered in thirteen sections of a class entitled "The Psychology of Child Development." These classes were conducted in eleven different communities under the auspices of the Extension Service of a large midwestern university. Each class met once a week for sixteen weeks. Ninety per cent of the students in these classes were school teachers. Eighty-one per cent were graduate students and seventy-four per cent were parents.

Data were obtained through questionnaires filled out at the end of the course and a test and questionnaire administered six months after the course was completed. Neither the questionnaires nor the tests were signed by the students and the grades were established before the instructor saw any of the questionnaires.

In the instructor-centered sections class time was distributed as follows: informal lectures, nineteen hours; films, five hours; small discussion groups, five hours; examinations, three hours. Students were expected to read the text and pass the examinations. The instructor attempted to clarify and expand the material in the text. This was done through informal lectures. Students were

encouraged to ask questions, give opinions, etc. at any time during the class. The instructor halted questions when they strayed outside of the unit covered during that period. The instructor attempted to be informal, friendly, and helpful in meeting student needs. At various times people in the community were brought in as resource people. This was often done at the suggestion of someone in the class. The procedure for evaluation was handled through two comprehensive examinations, both of which consisted of approximately sixty per cent objective type questions and forty per cent essay.

In the student-centered sections class time was distributed as follows: films, three and one-half hours; class discussion, twenty-eight and one-half hours. The students' responsibilities were to read the text and write between fifteen and twenty personal reactions to either items within the text or within the class discussions. This requirement of writing personal reactions was patterned after Nathaniel Kantor's method (*4*). Students were encouraged to put down their real feelings about class content or method. Actually the only assignments which were returned to the students as unsuitable were those which consisted of either irrelevant material or else an outline of subject matter rather than a personal reaction. Whenever possible in the student-centered classes the entire class sat around a group of tables with the instructor moving from one spot to another from week to week. When a procedural matter was under discussion ("When shall we see films?; How can we best use them?; Should we have a secretary to keep notes of class discussions?" etc.) the instructor attempted to act as discussion leader if no one else assumed that responsibility. During class discussions of the course content, the instructor seldom attempted to stimulate discussion, practically never made evaluatory comments concerning contributions, and during the first half of the course made a strong attempt to refrain from personal opinions on the items under discussion. At no time did the instructor call on a student. If asked for a personal opinion he tried to give a straightforward and honest answer. The instructor felt that the student-centered situation could most effectively promote significant learning by reducing to a minimum the individual threat to the learner, thus encouraging a more differentiated view of the subject matter. A rather common example of this within these classes was the feeling by some individ-

ual teachers that "some kids don't respect you unless you show that you can get tough with them." The holder of such an opinion might believe that others in the class, including the instructor, would look down upon him for holding such an opinion. Such feelings threaten his status with the group. This threat, and perhaps the guilt feelings attached to holding such an opinion, could very effectively reduce the ability of this person to accept other points of view. If he feels threatened by the holding of such an opinion, he must spend his time deciding how to defend himself rather than checking the validity of the opinion itself. The instructor attempted to reduce this individual threat by: (1) accepting all contributions as worthy and sincere; (2) paying full attention to each speaker, not waiting impatiently for him to finish so that the instructor could make his own comment; (3) pointing out that the teacher's rôle is one of the most difficult to play and that because of a difference in our background, etc., each teacher's approach to the classroom may be valuably different. In addition the instructor attempted to link the various comments together—that is, he attempted to point out the connections between individual speaker's contributions. At the end of each session the instructor made no attempt to provide closure. No final answers were given.

Classes smaller than twenty-five were given their choice as to learning method. The instructor presented an outline of the possibilities and requirements of the instructor-centered and student-centered classes and asked the students to discuss this in either small groups or as a whole and reach a decision while the instructor left the room. The four instructor-centered classes ranged in size from twenty-four to fifty-three; the mean class size was thirty-six. The nine student-centered classes ranged in size from eleven to twenty-four; the mean class size was eighteen. The instructor-centered classes were generally held in large industrial cities. These areas pay higher teacher salaries and may get better educated and more intelligent teachers than the smaller, more rural communities. There was a slightly higher percentage of graduate students in the instructor-centered classes. An Otis Mental Abilities Test was administered to two student-centered classes and one instructor-centered class. The latter scored slightly higher than both of the student-centered classes. In both learning situations the instructor attempted to be kind and understanding.

He made no attempt to vary either his own personality or the social climate. Verbal freedom was encouraged in both situations.

Six months after the course was completed a questionnaire and test were mailed to each student. Fifty-three students who had taken part in an instructor-centered course responded. In the six-month interval since the end of the course these fifty-three students had taken a total of thirty-eight courses in education or psychology. Seventy-four students who had taken part in a student-centered class responded. In the same interval they had taken a total of twenty-one classes in education or psychology. $\chi^2 = 25$, sig. < 0.01.

RESULTS AND DISCUSSION

On the basis of formal study in the six-month interval between the completion of the class and the examination, members of the instructor-centered sections should have demonstrated superior scores on the examination. This examination consisted of items previously used in either the mid-semester or final examination in the instructor-centered classes. Therefore the members of these sections were familiar with the items, type of test, etc. Members of the student-centered sections had not taken any tests and therefore found themselves faced with a new situation.

On the basis of four counts—(1) more intervening courses in education and psychology, (2) slightly higher proportion of graduate students, (3) slightly higher mean I.Q., (4) familiarity with test style and test items—it might be expected that members of the instructor-centered classes would achieve higher scores on the test. Actually no significant difference was found. Students who had participated in the instructor-centered classes achieved a mean raw score of 28.41 ($N = 53$). Former members of the student-centered classes achieved a mean score of 28.84 ($N = 74$).

Inasmuch as the instructor's major goal was to help the students use class subject matter as a guide for future professional behavior, the feelings of the student, i.e., his enthusiasm for the class and the subject matter, would seem to be important. Questionnaire results at the close of the course indicated some significant differences in the perceptions of the students in the instructor-centered and student-centered sections.

(1) Compared to other graduate courses in this general area, I

have learned:

	Much Less	*Less*	*About the same*	*More*	*Much more*	
Instructor-centered	4	21	68	23	15	(N = 121)
Student-centered	2	8	59	53	12	(N = 141)

$\chi^2 = 21.20$ sig. <0.01.

(2) As far as learning material which will be of practical use in my life, compared to other courses in this general area, I have learned:

	Much less	*Less*	*About the same*	*More*	*Much more*	
Instructor-centered	3	14	58	36	13	(N = 124)
Student-centered	0	5	45	68	23	(N = 141)

$\chi^2 = 20.50$ sig. < 0.01.

(3) Have you had any changes in attitude (not knowledge) as a result of this course?

	Yes	*No*	
Instructor-centered	49	36	(N = 85)
Student-centered	98	26	(N = 125)

$\chi^2 = 11.00$ sig. <0.01.

(4) Compared to other classes, this class has been:

	Very boring	*Dull*	*The same*	*More interesting*	*Much more interesting*	
Instructor-centered	3	18	60	36	5	(N = 122)
Student-centered	0	12	23	79	31	(N = 145)

$\chi^2 = 54.00$ sig. <0.01.

It is quite apparent that from the standpoint of the students, the student-centered courses were, on the whole, better and more interesting learning experiences.

It was not possible to obtain an objective measure of professional changes which could be ascribed to course participation. At the end of the course students estimated the amount of behavioral change that would result from class participation.

(5) How much change in your behavior will result from this course? (Guess)

	None	*Little*	*Pretty much*	*A great deal*	
Instructor-centered	7	70	34	12	(N = 123)
Student-centered	1	55	69	18	(N = 143)

$\chi^2 = 17.90$ sig. < 0.01.

Once more we have a significant difference in favor of the student-centered classes. Six months after completing the course students were asked if any behavioral changes had actually taken place.

(6) Have there been any changes in your own behavior which you could attribute to the course in "The Psychology of Child Development?"

	Yes	*No*
Instructor-centered	38	16
Student-centered	68	8

$\chi^2 = 7.63$ sig. <0.01.

It can be seen that the earlier perceptions of the students held up pretty well.

The questionnaire also indicated that there was no significant difference between the two groups in the amount of studying done in the area of child psychology or the use of the text as a reference in the six-month period following the class. It is probably safe to assume that graduate courses in education are designed to augment the professional equipment of the students enrolled in such courses. There is, however, a lack of evidence that most courses in education do help the students become better teachers. When students are asked whether they attempt to apply in their own classrooms the theories dealt with in these courses they say, "That stuff is too theoretical. It would never work in our school," or, "I get enthused in class but when I try it out it isn't practical." If these teachers are to actually apply the subject matter dealt with in such courses it is probable that two changes must be made. First, a change in attitude is necessary. Second, the subject matter must be learned more thoroughly.

It is easy to assume that a change in knowledge will result in behavioral changes. Too often this is not the case. If behavioral changes are to occur they must be preceded by changes in attitude as well as knowledge.

If the instructor's goal is an increase of knowledge, the method used to bring about such an increase may well be of little import. If the instructor's goal is to change the professional behavior of his students (so that their classroom performance is guided more by scientific principles and less by animism) the method used may be of major importance.

Take, for example, the use of verbal controls in the classroom. Evidence indicates that the type of controls used by the classroom teacher is correlated with the teacher's effectiveness (*10*). Informal, unpublished studies suggest that teachers of all ages can make changes in the type of controls they use in the classroom. In each instructor-centered class the instructor included in his lecture a suggestion for a simple method of self-training for teacher improvement in the use of verbal controls. A few weeks later the instructor asked the class if anyone had attempted to use this method of self-training. Approximately two per cent of the students said that they had done so. If the matter of verbal controls was brought up in a student-centered class, the instructor mentioned this same method for self-training. In subsequent classes a student generally brings out the fact that he tried this method. Other students relate their use of the method and the discussion usually indicates that between eight and twelve per cent of the students have actually attempted self-training in the use of their classroom verbal controls. On an examination, students in the instructor-centered classes easily recall that the type of verbal controls used in the classroom is important. They have, then, the knowledge necessary to make an improvement in their teaching. However, they seldom apply this knowledge in their own classrooms. What is needed is not the simple acquisition of knowledge but a change in behavior. This is generally not brought about through reading and listening.

There is evidence that such changes in behavior can be achieved through group discussion and decision. Lewin's data (*8*) suggest that decisions involving behavior changes are more easily effected by group discussion than by listening to a lecture. A follow-up in one case showed behavior changes as a result of group discussion in thirty-two per cent of the members. The lecture method produced behavior changes in only three per cent of the listeners. Subsequent variations of this experiment followed the same pattern. The change after lectures was in all cases smaller than the change after group discussion. We have, therefore, an indication that the lecture method of teaching, although it may be an efficient vehicle for the transfer of knowledge, may not be an efficient method to produce behavioral change in the student.

It is difficult to estimate the affective state of the students in these thirteen classes. The questionnaire obtained at the final

class meeting indicated that the students generally felt free to participate within both class frameworks.

(7) Did you feel free to make contributions to the class discussion?

	Yes	*No*	
Instructor-centered	78	9	(N = 87)
Student-centered	113	11	(N = 124)

$\chi^2 = 0.11$ no significant difference.

The increased interaction in the student-centered classes encouraged greater friendliness and self-involvement. Many of the one hundred and fifty students enrolled in the instructor-centered classes took no part in group discussions. For some of them their classroom cocoon was undoubtedly comfortable. Without any prodding on the part of the instructor each of the one hundred and forty students in the student-centered classes made some contribution.

No official attendance was taken but more absences were observed in the instructor-centered courses. This may have been a reflection of the higher student interest in the student-centered groups. This interest, in turn, may have been a function of the novelty of this learning situation.

Six of the nine student-centered classes seemed to develop rather high morale. During the closing meetings four of these made an effort to keep the group together through the scheduling of another class in the same community. Two of these groups became so involved in the subject matter that they occasionally extended the class time beyond the closing hour even though the instructor had to leave to teach another class.

SUGGESTIONS FOR THE USE OF A STUDENT-CENTERED METHOD

In a graduate course in education students are usually very curious about the workings of a student-centered method. They elect such a method, however, with a few misgivings. If the instructor limits his contributions to helping the class coördinate the discussion, the first few meetings are likely to go rather well. Students who enjoy talking are given an adequate chance to do so and students who are more reticent are pleased to learn that they will not be forced to participate. After five or six hours of discussion

the more dominant ego needs have been met and a drop in interest and amount of participation may be encountered. It is quite likely that the first few discussions have been interesting only because the people have been interesting. The group has not been very skillful in exploiting the subject matter. It is probably wise for the instructor to prepare the class for this letdown in morale in advance of its occurrence. After four or five hours of discussion have taken place the instructor might make a statement like this. "It seems to me that class has gone rather well. It would not be at all surprising if we had a letdown in interest or morale. If this does occur it probably will not last very long. Let's see it through. The class decided to elect this method of learning and after getting to know you better I feel that it was a wise choice. If, however, midway through the semester you would like to change to another method of learning it will be quite possible to make such a change." Normally, after a few uninteresting meetings the group will grow in discussion skill and will learn to profit academically and emotionally from this new classroom procedure.

Periodical evaluation ought to be very helpful. This can be done through either post meeting reaction forms (probably unsigned) or a general discussion of "how are we doing as a group?" It is probably best to vary evaluatory methods. Sometimes it is helpful for the instructor to ask someone to observe the problem-solving procedures used by the group and to give a short report to the group at the close of the meeting. Critical or even hostile people can be quite useful to the group in this capacity. After such an experience the evaluator often becomes a much more constructive group member.

In a student-centered learning situation the text materials which form the basis for discussion ought to be relatively free of dogmatism. If one text is used it is desirable that it present with some clarity a variety of data and a minimum of value judgments. This encourages students who tend to accept the truth of all printed matter with little or no thought to form their own opinions on the basis of the data and the class discussions.

It seems fairly clear that class size and physical conditions are limiting factors to the usefulness of the student-centered method. In the course of this experiment six of the student-centered classes were seated at a rectangular or square arrangement of tables. Three of the classes were seated in chairs in a circular arrangement.

The instructor felt that these three classes were slower than average in the development of group spirit and a relaxed, free atmosphere. The best possible arrangement would probably be a circular table of such a size that members would be sitting close together but not cramped. The larger the group the more important it becomes that the meeting be held in a room with good acoustics and a minimum of outside noise. Under normal conditions, as the size of the group increases beyond approximately twenty, the forces inhibiting free communication also increase. In an informal discussion group of thirty people, if some members are to contribute they apparently need considerably more motivation to speak than they would need as members of a group of twenty. In the present experiment no classes of more than twenty-five used the student-centered method. This was based on an arbitrary judgment of the experimenter. The data yield no relationship between class size and student performance or satisfaction within either of the learning situations. Under optimum physical conditions, and with a wise, warm and understanding teacher, it is possible that larger classes might profit from such an approach.

One of the difficulties faced by those who wish to experiment with classroom learning methods is the impossibility of duplicating classroom situations. Not only is each class different by virtue of different student personalities, but each instructor's needs are different and the restrictions imposed by the institution are quite dissimilar. Previous research concerning classroom learning methods has largely ignored the ego-needs of the instructor (*9, 11*). These needs play a vital rôle in the results obtained. On the basis of this research, the experimenter suggests that it is wise for the instructor to be as honest in regard to his ego needs as the situation will allow. In a truly student-centered classroom everyone in the room shares responsibility for the content and the methods used to handle that content. If the instructor feels that he would be derelict in his professional duties if certain aspects of the course content are not adequately dealt with in class, then he ought to express his feelings at the outset. Such a straightforward approach will be appreciated by the students and will interfere less with their self-motivation than a more circuitous method.

The stimuli for instructor self-gratification are quite different in the student-centered classroom. The glow of self-appreciation following the presentation of a well-planned lecture may be sorely

missed. There are, however, multiple satisfactions to the instructor using such a method. Many of these stem from the new insights gained and expressed by students. One member of a student-centered class in a highly rural section made only one contribution during the first five weeks of class. She volunteered the information that she was late because the hired man was sick and she felt that it was more important to "milk those poor cows than to get to class on time." During the sixth class session the topic was "The Affective Life of the Child." Midway through the meeting she said: "I think what we are trying to say is that learning is like a radio. Radio waves are going through the air all of the time but the radio doesn't pick them up until the switch is turned on. Things are happening around us all the time, too, but we don't learn until the switch is turned on. Our emotions are that switch." This was rewarding.

SUMMARY

This study was designed to investigate the effectiveness of two learning situations in achieving the following goals: (1) to help students learn as much as possible about present knowledge in the area of child psychology, and (2) to help students use this knowledge to guide their own behavior.

An instructor-centered approach was used in four sections and a student-centered method was used in nine sections. Ninety per cent of the students in these classes were school teachers, eighty-one per cent were graduate students, and seventy-four per cent were parents. Data were obtained through questionnaires administered at the end of the course and through a test and questionnaire administered six months after the course was completed.

On the basis of four counts—(1) more intervening courses in education and psychology, (2) slightly higher proportion of graduate students, (3) slightly higher mean I.Q., (4) familiarity with test style and test items—it might be expected that members of the instructor-centered classes would achieve higher scores on the test. Actually no significant difference was found. Students who had participated in the instructor-centered classes achieved a mean raw score of 28.41 ($N = 53$). Former students in the student-centered classes achieved a mean raw score of 28.84 ($N = 74$).

At the close of the course, members of the student-centered sections estimated that they had learned more, that what they

had learned would be of more practical use to them, that more attitude change had taken place, that more behavioral change would take place as a result of the course, and that the class had been more interesting. All differences between the instructor-centered and student-centered groups were significant at greater than the 0.01 level and favored the student-centered sections.

During the six-month period following the class there was no significant difference between the two groups in the amount of studying done in the area of child psychology or the use of the text as a reference. Six months after the completion of the course, the student-centered sections estimated more behavior change as a result of the course than did the instructor-centered sections (significant at better than 0.01).

REFERENCES

1) M. J. Asch, "Nondirective Teaching in Psychology; An Experimental Study," *Psychology Monograph*, 65: 4, 1951.

2) E. W. Bovard, Jr., "The Psychology of Classroom Interaction," *Journal of Educational Research*, 45: 215–224, 1951.

3) E. W. Bovard, Jr., "Clinical Insight as a Function of Group Process," *Journal of Abnormal and Social Psychology*, 47: 534–539, 1952.

4) N. Kantor, *The Dynamics of Learning*. Buffalo, Foster and Stewart, 1946.

5) V. Faw, "A Psychotherapeutic Method of Teaching Psychology," *American Psychologist*, 4: 104–109, 1949.

6) L. M. Gibb and J. R. Gibb, "The Effects of the Use of 'Participative Action' Groups in a Course in General Psychology," *American Psychologist*, 7: 247, (abstract), 1952.

7) H. Kelley and A. Pepitone, "An Evaluation of a College Course in Human Relations," *Journal of Educational Psychology*, 43: 193–209, 1952.

8) Kurt Lewin, "Group Decision and Social Change," *Readings in Social Psychology*, pp. 459–473. New York, Henry Holt and Co., 1952.

9) W. J. McKeachie, "Student-Centered versus Instructor-Centered Instruction." *Journal of Educational Research*, 45: 143–150, 1954.

10) W. C. Olson and M. M. Wilkenson, "Teacher Personality as Revealed by the Amount and Kind of Verbal Direction Used in Behavior Control," *Educational Administration and Supervision*, 24: 31–39, 1938.

11) H. Ruja, "Experimenting with Discussion in College Teaching: a Survey of Recent Research," *Educational Administration and Supervision*, 39: 321–342, 1953.

THEORY DEVELOPMENT AND THE STUDY OF TEACHER BEHAVIOR

DAVID G. RYANS

University of California at Los Angeles

THE RÔLE OF THEORY

Teacher behaviors, their nature, genesis, and cultivation, their identification and their evaluation, constitute the core concern not only of teacher training institutions and school systems, but also of a society at large that depends on teachers to a very great extent for the propagation of accumulated knowledge and cultural values. Considerable opinion is expressed and not a few studies are undertaken in an effort to better understand teacher behavior. (*1*, *2*, *4*, *5*, *8*, *10*, *11*, *13*, *14*.) Relatively little attention, however, appears to have been directed toward the organization or systematization of what is known about teacher behavior or the assumptions on which the study of teacher behavior is based.

In the meantime, strong arguments have been presented by social and behavioral scientists suggesting that steps toward some sort of a theoretical formulation have considerable utility value; that systematic theories are highly desirable, particularly for guidance, if maximum productivity of research in a field is to be attained. Proponents of this view point out that advanced understanding and usable knowledge frequently have been observed to demonstrate marked increase as the study of problems in an area progressed from exploratory, "catch-as-catch-can," investigations to selective observation guided by hypotheses derived from systematic theory and employing empirical tests to determine the place (if any) of such hypotheses in the basic theory.

Essentially, this approach represents a rapprochement of formal, deductive logic and a modern science that demands the inclusion of pragmatic considerations in judging the admissibility of facts to a body of knowledge. Such axiomatic, or hypothetico-deductive, theory is rather popular at the present time.

Another sort of theorizing attempts to be rigorously inductive and culminates in what may be simply described as an organized body of empirically obtained facts. This kind of theory employs operational definitions extensively and emphasizes reasoning that

proceeds from the observed characteristics of sample data to generalizations applicable to larger homogeneous classes. Facts are admissible to such a theory if (1) they are based on fair samples, (2) they satisfy pragmatic criteria, and (3) they fit into a pattern with other known facts in the same area. Hypotheses may be derived from the pattern of accumulated fact (and to this extent the approach is also deductive). Inductively generated theories, or systematizations, are descriptions rather than explanations.

Researchers and practitioners appear to have found little need for either kind of theory development in approaching the problem of teacher behavior. A recent statement attributed the lack of productive research on teacher effectiveness, in part, to such a neglect of theory, declaring:

> . . . the present condition of research of teacher effectiveness holds little promise of yielding results commensurate with the needs of American education. This condition has two significant characteristics: disorganization, and lack of orientation to other behavioral sciences. By disorganization, we mean the condition in which, at present, research too often proceeds without explicit theoretical framework, in intellectual disarray, to the testing of myriads of arbitrary, unrationalized hypotheses. The studies too often interact little with each other, do not fall into place within any scheme, and hence add little to the understanding of the teaching process. The simple fact of the matter is that, after 40 years of research on teacher effectiveness during which a vast number of studies have been carried out, one can point to few outcomes that a superintendent of schools can safely employ in hiring a teacher or granting him tenure, that an agency can employ in certifying teachers, or that a teacher-education faculty can employ in planning or improving teacher-education program. (*2:* p. 657.)

Something more should be said, perhaps, of the essential characteristics of theories and the values that are claimed for them. Rose, in his volume of essays on theory in the social sciences, states rather well the generally accepted concept of theory viewed from the standpoint of formal logic with appropriate adaptations to make place for a pragmatic criterion of validity. Rose writes that theory is:

> . . . an integrated body of definitions, assumptions, and general propositions covering a given subject matter from which a comprehensive and consistent set of specific and testable hypotheses can be deduced logically. The hypotheses must take the form "If a, then b, holding constant c, d, e, . . ." or some equivalent of this form, and thus permit of causal explanation and prediction. (*12:* p. 3.)

Rose goes on to say:

> A *good* theory is one which 1) has its definitions, assumptions, and general propositions consistent, insofar as possible, with previous research findings and with careful, although perhaps not systematic, observations; 2) has a minimum number of definitions, assumptions, and general propositions; 3) has its deduced hypotheses in readily testable form; and, crucially, 4) gets its deduced hypotheses verified by proper scientific methods. (*12:* p. 3.)

Just how does such a theoretical framework, in the view of its proponents, really serve in leading to a more complete understanding of a field? Is all of this merely a scientist's high-sounding words —"gobbledygook" (exposition that is obscured by excessive use of technical terminology and involved phrasing)? Or does it have some real utility? It would seem that such a theory frequently does have utility; that it does provide certain advantages to workers in a field. Thus, it may serve to give proper organization to our thinking about, and study of, a problem. It may help us to specify problems and formulate them in a manner that permits the seeking of direct answers. It may call attention to the probable futility of pursuing certain contradictory hypotheses, hypotheses that might, in the absence of the theoretical setting, seem plausible. It may help us to locate strategic, or crucial, problems. All of these suggestions may be summarized in what many scientists and philosophers feel to be a significant value of theory, the economy of thought and labor that may be affected when investigation is conducted in a theoretical framework.

Rose, to whom reference was made earlier, says of the utility of theory:

> . . . 1) It is a guide to the formation of hypotheses and trains investigators to look for facts which may ordinarily not be readily apparent. 2) It permits research to be cumulative; that is, it allows the conclusions of older studies to gain support from new research and allows older studies to provide some of the data for new research. 3) It indicates what studies are crucial; that is, it provides one guide for the selection of research problems from among the infinitely large number of possible hypotheses. 4) It permits research to proceed systematically and allows conclusions to take shorthand form so they are readily communicable. (*12:* p. 4.)

SOME FURTHER CHARACTERISTICS OF THEORIES

All theories, whether of (1) an axiomatic, hypothetico-deductive type or (2) an empirical, inductive sort, must begin with certain

assumptions and perhaps some definitions. In the case of inductive theory, such assumptions are very largely expressions of faith in the inductive method—in the consistency and observability of nature; and the definitions are predominantly of the operational kind. In axiomatic theory the assumptions are likely to be the working premises, or general propositions, from which the theory proceeds to noncontradictory hypotheses (postulates and theorems derived from the assumptions or general propositions), that may be tested to determine their probable validity. Inductive theory plunges at once into observation, stressing classification and cross-classification, on the one hand, and the experimental testing of sporadic hypotheses, on the other. The resulting specific descriptions of properties and of functional relationships lead, with confirming replications, to generalizations in the form of inductive principles and laws, which in turn may be systematized and organized into coherent theory. The ultimate goal of theory, either axiomatic or inductive, is prediction.

The assumptions of any theory are simply judgments or propositions that are accepted, or taken for granted. Thus, depending on the kind of theorizing we are attempting, the assumptions may be based on facts held to be axiomatic; or, they may be propositions deductively derived from either axiomatic or previously confirmed premises; or, they may be based directly on previously obtained and admitted scientific evidence (i.e., evidence accepted as valid, or admissible, because it already has passed satisfactory tests); or, again, the assumptions may even be concepts or propositions that are accepted (even though they are not capable of immediate confirmation in any manner), because they are necessary for the investigation of phenomena in the area with which the theory deals.

In the case of hypothetico-deductive theory, hypotheses are derivable from the assumptions or basic propositions, growing out of questions, problems, or needs suggested by the basic assumptions with which the theory begins. In inductive theory, hypotheses are suggested by experience—previous observations, or empirically obtained evidence. In either case, hypotheses are essentially creations of the imagination of their inventors or originators. In an unrefined and inexact form, they begin to appear as soon as the person develops any definite conception of a thing or an event (e.g., "I think a teacher needs a course in *X* subject matter to per-

TABLE I.—THEORETICAL FRAMEWORK

Type I. Axiomatic, Mathematico-Deductive, Hypothetico-Deductive Theory	Type II. Inductive, Empirical Theory
A. Assumptions Definitions	A. Assumptions Definitions (especially "operational")
B. General Propositions	B. Observation Classification and cross-classification Sporadic hypotheses and their experimental testing
C. Hypotheses: postulates and theorems re properties and functional relationships	C. Specified descriptions of properties and functional relationships (generalizable to inductive principles and laws)
D. Observation: testing of postulates and theorems by classification and experimentation	D. Systematization, organization of descriptive principles into coherent pattern (possibly leading to General Propositions)
E. Extension and revision of theory in light of observation and of further deductive analysis	E. Extension and revision of principles, and of theory, as required by further classification and experiment
F. Prediction	F. Prediction

form his teaching responsibilities adequately." "I think older teachers are more authoritarian than younger ones."). But all such ruminations probably should not be thought of as hypotheses—at least, not in the sense in which the term hypothesis usually is employed in science. There are certain generally agreed-upon restrictions that commonly are applied. Thus, a suitable hypothesis must be in an exact form; must provide a specific answer to a specific question. And it must not be inconsistent with what is known; it must not be contradictory to the basic assumptions and propositions of applicable theory, or with other hypotheses for which there is extensive supporting evidence. And, again, a suitable hypothesis must be stated in a way that it is capable of being confirmed or refuted at some agreed-upon level of confidence; the consequences that will follow if it is probably true, or probably false, must be describable and such consequences must be capable of observation and recording.

SOME STARTING POINTS FOR A THEORY OF TEACHER BEHAVIOR

Thus far our discussion has been concerned with comments on the nature of theory. Now how can some of these concepts relating to theoretical framework be applied to teacher behavior? What, for example, might be some of the assumptions and definitions that apply to teacher behavior? Those that will be suggested do not constitute a complete set of all assumptions required for a theory of teacher behavior. Nor is any particular claim made at this point for theoretical rigor. But if we do wish to think in terms of theory, a starting point is necessary regardless of how crude and tentative such a beginning may be. Trial and error, criticism, and research may provide the necessary clarification, revision, and extension leading to a generally acceptable theory of teacher behavior.

A definition. As a point of embarkation, a definition of teacher behavior will be stated. It is suggested that for our purposes teacher behavior may be defined simply as the behavior, or activities, of persons as they go about doing whatever is required of teachers, particularly those activities that are concerned with the guidance or direction of the learning of others.

There are several important implications in this rather straightforward operational definition.

(A) One implication is that teacher behavior is social behavior—that, in addition to the teacher, there must be learners, or pupils, who presumably are affected by the behavior of the teacher. It should also be noted that teacher behavior-pupil behavior relationships may be of a reciprocal nature; that not only do teachers affect pupil behavior, but pupils may affect teacher behavior, as well.

(B) Another implication of the definition of teacher behavior is that there is nothing inherently good or bad in a given teacher behavior, or set of behaviors, but, instead, that teacher behavior is good or bad, right or wrong, effective or ineffective, only to the extent that the behavior conforms or fails to conform to the particular value system, or set of objectives, defining (1) the activities expected of a teacher in a given community or culture, and (2) particularly the kinds of learnings (attainments) and methods of teaching to bring about these learnings, approved by the particular culture (*11*, *14*).

Assumption I. If we were to state the basic assumption to be

made in setting out to describe teacher behavior, we might expect it to bear some resemblance to formulations made for similar purposes in connection with learning theory and personality theory. Indeed, in any and all behavior theory, some expression of faith in the reliability, or consistency of behavior is required. In our case the assumption may be summarized in the proposition that teacher behavior is a resultant of (a) certain situational factors and (b) certain organismic factors, and their interaction; or, simply that teacher behavior is a function of certain environmental influences and characteristics of the individual teacher. Thus:

$$\text{(A)} \quad r_a = f[(S_{1_a} M_1{}^o E_1{}^o H_1{}^o) + (S_{2_a} M_2{}^o E_2{}^o H_2{}^o) + \cdots (S_{n_a} M^o n E^o n H^o n)]$$

where:

r_a = teacher behavior r in situation a
S = situational indices
M^o = motivational organismic indices
E^o = experience organismic indices
H^o = genetic organismic indices

$$\text{(B)} \quad R_t = ar_a + br_b \cdots nr_n + e$$

where:

R_t = over-all teacher behavior
a = weight
r_n = z score of teacher behavior n
e = z score of error component

The present treatment is intentionally limited to consideration of teacher behavior per se. Since, however, teacher behavior implies pupil behavior (as previously noted), we might also observe that pupil behavior could be thought of as being similarly influenced by a number of variables, one of which would be teacher behavior R_t. Thus:

$$\text{(C)} \quad r_a + f[(S_{1_a} M_1{}^o E_1{}^o H_1{}^o) + \cdots (S_{n_a} M^o n E^o n H^o n)]$$

where:

S represents situational indices such as textbooks, peer behavior, parent behavior, teacher behavior, etc.

and

$$\text{(D)} \quad R_p = ar_a + br_b \cdots nr_n + e$$

where: R_p = pupil behavior

What these rather rigid appearing equations say is simply that behavior occurs when an organism is stimulated and if the variables and mathematical functions are known, the equation may be solved for the unknown (r or R).

Just how these various situational and organismic conditions interact and what takes place in the teacher's nervous system as they interact (how, as some theorists would put it, the energy input-output transfer takes place) certainly is not known and we are completely incapable of describing the process except in terms of inferences based upon observable[1] inputs and observable responses of the teacher. The fact that little is known about how such things take place does not mean that persons interested in behavior theory have not been actively concerned about the problem. Certain groups of theorists have been both active and ingenious. One such group, which incidentally is interested in speculating upon the generality of behavior theory whether the systems involved are atoms, viruses, cells, individual persons, society, solar systems, or what not, view the organism, the teacher in our case, as an open system (i.e., a bounded region in space-time), possessed with negative feedback which provides for the distribution of information to sub-systems to keep them in orderly balance.

Other theorists have been particularly concerned with how individuals in a social environment (e.g., teacher-pupil, teacher-teacher, etc.) interact and condition one-another's behavior, not only in an immediate situation but also in future situations, as a result of the integration of response-produced stimuli into the total stimulus pattern. Sears (*15*) has suggested, for example, the necessary expansion of the basic monadic unit of behavior, which various learning theories have employed—and in terms of which the foregoing equations (A), (B), (C), and (D) are presented—into a diadic one which describes the combined actions of two or more individuals. That a diadic unit is essential if the relationships between people are to be taken into account in theory should impress the teacher behavior theorist as an entirely reasonable proposition.

The diadic approach strikes at the heart of the teacher-pupil relationship problem. Therefore, it is important to take into account extension of the paradigm represented by our equation (A)

[1] Observable, as here used, refers to that which may be perceived either by another person or by the experiencing individual, or which may be recorded by some instrument.

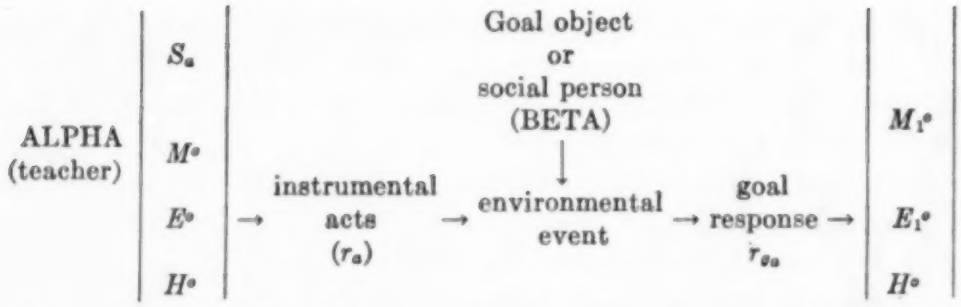

FIGURE 1. The monadic instigation-action sequence. After Sears (*15*)

to include the significant implications of Sears' model for the diadic behavior sequence. The models which follow conceptualize teacher behavior first in terms of the monadic and then the diadic behavior sequence. Alpha may be presumed to be the teacher, and Beta the pupil. The symbols S_a, M^o, E^o, H^o, and r_a have the same meaning as in the earlier equations. $M_1{}^o$, and $E_1{}^o$ are employed to represent the modified motivational and experience structures of the teacher resulting from a reinforcement brought about by completion of the behavior sequence.

We are proposing, then, that the basic assumption of teacher behavior theory may be summarized by the equations and the models just presented.

Now growing out of the basic assumption that teacher behavior is a function of the conditions under which it occurs are a number of implications or sub-assumptions, which follow:

(A) An implication of this assumption is, of course, that teacher behavior (and social behavior, with which education deals), is characterized by some uniformity; that, as Mill put it: ". . . there

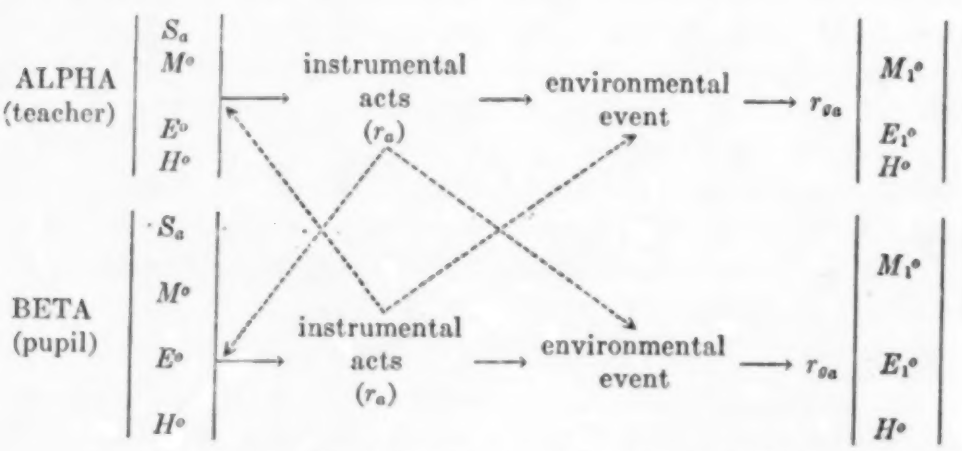

FIGURE 2. The diadic sequence. After Sears (*15*)

are such things in nature as parallel cases, that what happens once will, under sufficient degree of similarity of circumstances, happen again" (*9:* p. 223). We are stating that teacher behavior is, within limits, reliable, consistent, or stable.

(B) Another implication of this basic assumption (perhaps it is so fundamental to any or all scientific theory that it is unnecessary to state it explicitly with respect to teacher behavior), is expressed by Keynes' Postulate of Limited Independent Qualities, which states that: ". . . objects in a field over which our generalizations extend, do not have an infinite number of independent qualities; . . . their characteristics, however numerous, cohere together in groups of invariable connections, which are finite in number" (*7:* p. 256). This is to say that the number of responses the individual teacher is capable of making, and the number of stimulus situations and organismic variables that may affect a teacher's behavior, are limited. The assumption is an important one if the equations noted above are to hold, and, of course, if we hope to predict teacher behavior. It presents the researcher with a "tolerable" problem (as a student recently put it).

(C) In taking cognizance of this basic proposition of teacher behavior, it should be noted that human behavior, characterized as it is by variability (only approximate uniformity, or consistency), always must be considered in light of probability rather than rigid, one-to-one, invariable cause-effect relationships. The error component (unreliability) resulting from behavior sampling inevitably will be present in any assessment we make of stimulus conditions, of organismic conditions (genetic bases, past experience, motivation), and of the teacher behavior (the dependent variable, or criterion). Teacher behavior can be predicted only with varying degrees of probability; never with certainty. Therefore, r_a and R_t refer to probable behaviors rather than certain behaviors.

Assumption II. When we presume to study teacher behavior we also make the assumption that teacher behavior is observable; that its manifestations are of a tangible nature and may be identified objectively, either by direct observation of samples of behavior, or by approaches that provide correlative indices of teacher behaviors (e.g., approaches involving the assessment of pupil behavior, the use of tests of teacher abilities and knowledges, the use of interviews or inventories to elicit expression of teacher preferences, interests, beliefs, and attitudes, etc.).

Several implications (or perhaps subassumptions) may be noted here.

(A) One implication is that different teacher behaviors, or sets of teacher behaviors, are distinguishable from one another; that some behaviors possess characteristics in common, to the exclusion of other characteristics. Teacher behaviors can be distinguished under observation.

(B) A second subassumption is that the distinguishing characteristics of a teacher behavior provide the basis for a concept, abstracted from specific instances, which permits common understanding of the behavior and communication of the behavior to others.

(C) These implications lead to a third aspect of the assumption of the observability of teacher behavior, that teacher behaviors are classifiable, both qualitatively and quantitatively. (1) Teacher behaviors that are similar, that have common elements, may be classified in the same category—apart from other teacher behaviors that are dissimilar. (2) Within any given category teacher behaviors may be further assigned to subclasses that may be treated quantitatively (to which numbers may be attached). These quantitative subclasses may be of a present-absence type permitting enumeration, or counting, only; or, they may be subclasses characterized by continuity and varying as a metric (exemplified by ordinal subclasses at the lowest level of refinement and by equal-interval and equal-ratio subclasses at successively more refined levels).

Various qualitative classifications of teacher behaviors are possible. Thus, for example, we might think of teacher behaviors grouped broadly into such general categories as: (1) those involving instruction and relationships with pupils; (2) those involving relationships with the school, its organization, and its administration; and (3) those involving relationships with the community. Or, we might classify teacher behaviors more specifically, employing such categories as verbal aptitude, emotional stability, favorable attitude toward pupils, friendly behavior toward pupils, and the like, to provide the framework. For any specifically defined category we might conceivably assign quantitative subclasses.

(D) Still a fourth subassumption to the basic assumption of observability of teacher behavior might state that teacher behaviors may be revealed, or may be observed, either (1) by the representative sampling of teacher behaviors or (2) by signs, or

indicators, or correlates of the behavior under consideration. In sampling behavior, we assume that the performance of the individual during the behavior sample is proportional to the larger aspects, or universe, of his behavior. In judging behavior from signs or correlates, it is assumed that behavior can be inferred from known correlates of that behavior—from phenomena that have been found to accompany that behavior in the past.

So much for our articles of faith, or general assumptions regarding teacher behavior. With such assumptions stated (and assuming for the present the desirability of inductive theorizing), the teacher behavior theorist might turn to observation—to classification and hypothesis testing.

The number of descriptive classifications and specific hypotheses that might be generated with respect to teacher behavior is almost limitless. Probably no one would be interested in all such hypotheses even if it were possible to assemble them. But some classifications and some hypotheses are more relevant than others. And, moreover, some appear to be quite plausible. In fact, many can even be incorporated in the design of well-controlled study.

At this point it might be well to note that there are various approaches to the testing of hypotheses. Ideally, to test any hypothesis, we would like to design a projected experiment (the basic experimental design) with (1) experimental and control groups equated at the beginning of the investigation, (2) introduction of an experimental variable or treatment (the hypothesized causal factor) with the experimental group while withholding it from the control group, and (3) judgment of the validity of the hypothesis by comparison of the groups with respect to the dependent variable, or criterion behavior. This is the typical application (or, it may involve some more advanced and complex modification to permit analysis of interaction effects and such), of Mill's methods of agreement and difference. But this approach frequently is not possible in the study of teacher behavior. Instead, it usually is necessary to undertake what Chapin (*3*) has called "ex post facto" experiments, or perhaps the comparison of aspects of defined groups, or perhaps intercorrelational studies (classificatory investigation). Much of the research on teacher behavior must be of this nonexperimental (using "experiment" in the strictest sense) type.

No attempt will be made to state specific hypotheses about

teacher behavior that might be subjected to testing. Tests of a number of hypotheses have been attempted in connection with various researches (*10*), including the Teacher Characteristics Study (*14*). In the latter research, hypotheses referring to the interrelationships of estimates of teacher behaviors have been extensively tested. Similarly, hypotheses relative to the central tendencies and dispersions of behavior measurements of defined groups of teachers have been given attention, leading to comparisons of teachers with respect to such conditions as age, experience, geographic area of employment, type of college attended, youthful experiences, frequent activities, etc., on a selected set of teacher personal and social characteristics.

Throughout the country, many similar researches are being directed at the study of comparisons of, and interrelationships among, teacher behaviors. Perhaps we are reaching the point where patterns among the behavior components may be discerned—where the beginnings of organization of evidence into a theory of teacher behavior may be possible.

Such study and systematization (theory development), in combination with proper attention to specification of problems and careful design, may reasonably be expected to shed increasing light upon our understanding, first, of teacher behavior, and, second (and of ultimate importance), on how teacher behavior influences pupil behavior and how certain kinds of teacher-pupil relationships may be identified, predicted, and cultivated.

REFERENCES[2]

1) A. S. Barr, B. V. Bechdolt, W. W. Coxe, N. L. Gage, J. S. Orleans, H. H. Remmers, and D. G. Ryans, "Report of the Committee on the Criteria of Teacher Effectiveness," *Revue of Educational Research*, 22: 238–253, 1952.

2) A. S. Barr, B. V. Bechdolt, N. L. Gage, J. S. Orleans, C. R. Page, H. H. Remmers, and D. G. Ryans, "Second Report of the Committee on Criteria of Teacher Effectiveness," *Journal of Educational Research*, 46: 641–658, 1953.

3) F. S. Chapin, *Experimental Designs in Sociological Research*. New York, Harper & Bros., 1947.

[2] References 1, 2, 11, and 14 bear, to some extent, on theory of teacher behavior. References 5 and 10, in particular, provide bibliographies and reviews of studies relating to the evaluation of teacher behavior. It should be noted, also, that the June issues of the *Review of Educational Research* in the years 1934, 1937, 1940, 1943, 1946, 1949, 1952, and 1955 were devoted to "Teacher Personnel."

4) W. W. Charters and D. Waples, *The Commonwealth Teacher Training Study*. Chicago, University of Chicago Press, 1929.

5) S. J. Domas and D. V. Tiedeman, "Teacher Competence: An Annotated Bibliography," *Journal of Experimental Education*, 19: 101–218, 1950.

6) E. Greenwood, *Experimental Sociology*. New York, King's Crown Press, 1945.

7) J. M. Keynes, *A Treatise on Probability*. New York, Macmillan Co., 1921.

8) W. A. McCall, *Measurement of Teacher Merit*. Raleigh, N. C., State Department of Public Instruction, 1952.

9) J. S. Mill, *A System of Logic*. New York, Longmans, Green, Co., 1872.

10) J. E. Morsh, and E. W. Wilder, *Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900–1952*. San Antonio, Texas, Air Force Personnel and Training Research Center, 1954.

11) W. Rabinowitz and R. M. W. Travers, "Problems of Defining and Assessing Teacher Effectiveness," *Educational Theory*, 3: 212–219, 1953.

12) A. M. Rose, *Theory and Method in the Social Sciences*. Minneapolis, University of Minnesota Press, 1954.

13) D. G. Ryans, "Appraising Teacher Personnel," *Journal of Experimental Education*, 16: 1–30, 1947.

14) D. G. Ryans, "The Investigation of Teacher Characteristics," *The Educational Record*, 34: 370–396, 1953.

15) R. R. Sears, "A Theoretical Framework for Personality and Social Behavior," *The American Psychologist*, 6: 476–483, 1951.

CLASS AND SEX DIFFERENCES IN VERBAL FACILITY OF VERY BRIGHT CHILDREN

HELEN H. DAVIDSON and DOM BALDUCCI

The School of Education, The City College of New York

Social class status as a significant variable in the understanding of human behavior has preëmpted the position previously held by sex and I.Q. While there have been many studies demonstrating that differences in personality may in part be accounted for by differences in sex (*12*) and by differences in social class (*6*), the interaction between sex and social class status on human behavior has only recently been examined (7, 11).

It is the purpose of the present investigation to inquire further into the mutual influence of sex and class status on verbal facility. The ability to verbalize easily, to know and use appropriately many words, to understand the finer shades of meanings of words, is tremendously important in judging many facets of personality, from intellectual status to personality adjustment (*1*, *4*, *5*). Such verbal facility has been found in numerous investigations to be more advanced in girls (*10*), and likewise among children of the higher economic classes (*3*, *10*, *13*). It is therefore desirable to determine whether the verbal ability of children, all very bright but of disparate economic backgrounds, is also conditioned by social class, and whether it is the boys or the girls who are more affected by class position. Conceivably, bright children might be able to overcome the detrimental effects of low social status.

THE DATA

Data which might throw some light on these questions were available for ten boys and ten girls of high social class status and for ten boys and ten girls of low social class status.

These forty children were selected from a group of one hundred and two who were subjects in a previous investigation (*2*). Their ages ranged from nine years, three months, to fourteen years, one month, with an average age of eleven years, ten months for the boys and eleven years, six months for the girls. Their Stanford-Binet I.Q.'s ranged from 120 to 157, with an average I.Q. of 143 for the boys and 135 for the girls. Income of the family was used

as an index of social class. Information concerning income was obtained from school records and checked against other reliable sources of information. Family incomes of the low social class children ranged from "destitute" (families were on home relief) to thirty-nine hundred dollars per year; for the upper class children, family incomes ranged from six thousand dollars to over ten thousand dollars per year.[1] The high social class children were students at a private experimental school; the low social class children were students at a special experimental public school. The two groups were matched for age, I.Q., and kind of personality adjustment revealed by the Rorschach protocol.[2]

Verbal facility was measured in three ways using responses given to Cards I, III, VI, and X of the Rorschach test: (1) the total number of words used in the *initial* responses, exclusive of words such as "the," "a," "an," "and," "like," or "looks like"; (2) the number of uncommon words used;[3] and (3) the number of adjectives and adverbs. Product-moment correlations among the measures were: 0.58 (1 and 2); 0.53 (2 and 3); 0.71 (1 and 3). The size of these correlations justify the use of the three different modes of analysis.

Rorschach responses seemed particularly useful as measures of verbal facility since they are probably less determined by cultural experiences than are such data as vocabulary knowledge, the relating of experiences, telling a story, or conversation, as in play situations, frequently used in studies of language development. The latter approaches give a decided advantage to girls because of their greater opportunity to use language at home and in play (*9*), and to children of the upper class because of their greater experience with situations involving language (*8*). It may also be argued that since the population of the present study is somewhat older than samples used in previous investigations, the effects of family background have been offset by the four or five years of schooling these children have received.

---

[1] These data were gathered in 1940.

[2] All the Rorschachs were administered and interpreted by one examiner. For further details as to how personality was judged for adjustment, see reference number (*2*).

[3] A word was judged "uncommon" if it occurs eight or less times per million words in general reading matter, according to the tabulation in the *Teacher's Word Book of 30,000 Words* by E. L. Thorndike and I. Lorge, New York, Bureau of Publications, Teachers College, 1944.

RESULTS AND DISCUSSION

The null hypotheses to be tested by the application of analysis of variance to the data are: (1) there is no significant difference between the sexes; (2) there is no significant difference between the two social classes; and (3) the interaction between sex and social class status is not significant.

In Table I are given the means and standard deviations for the two social classes and for the sexes separately as well as for the total group, by sex and by social class. The analysis of variance results are presented in Table II.

The evidence indicates that there are no significant differences between the sexes or between the classes for any of the three measures of verbal facility used. The interaction between sex and class is likewise not significant. Although statistically the differences are not significant, the means of the three measures as observed in Table I reveal a consistent trend. The boys are superior to the girls and maintain their superiority in all three measures, regardless of social class position. On the other hand, when sex is disregarded, the high social class children are always superior to the low social class children. In all three verbal measures, the girls of the low social class are most disadvantaged.

It is possible that significant differences were not obtained because of the large individual variation present (note S.D.'s and error variance). The most nearly significant difference (between the two social classes in the number of uncommon words used) deserves special mention. Although very bright children seem able, to some

TABLE I.—MEANS AND STANDARD DEVIATIONS BY SEX AND SOCIAL CLASS FOR THREE MEASURES OF VERBAL FACILITY

	All Boys	All Girls	All Low Class	All High Class	Low Class Boys	High Class Boys	Low Class Girls	High Class Girls
Total number of words used	55.7* (26.1)	47.9 (23.3)	47.9 (26.3)	55.7 (24.1)	54.6 (27.8)	56.8 (24.2)	41.3 (22.9)	54.6 (21.9)
Number of uncommon words	2.8 (2.6)	2.4 (2.2)	1.95 (1.8)	3.25 (2.8)	2.0 (1.7)	3.6 (3.0)	1.9 (1.8)	2.9 (2.4)
Number of adjectives and adverbs	5.8 (4.5)	4.3 (3.4)	4.4 (3.8)	5.6 (4.3)	4.8 (4.3)	6.8 (4.5)	4.0 (3.1)	4.5 (3.6)

* The upper number is the mean; the number in parentheses is the standard deviation.

TABLE II.—ANALYSIS OF VARIANCE BETWEEN BOYS AND GIRLS AND BETWEEN LOW AND HIGH SOCIAL CLASSES FOR THE THREE MEASURES OF VERBAL FACILITY

	Source	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Total number words used	Sex	600.63	1	600.63	0.91
	Class	600.63	1	600.63	0.91
	Sex x Class	308.03	1	308.03	0.47
	Error	23600.50	36	655.57	—
	Total	25109.77	39	—	—
Number of uncommon words	Sex	1.60	1	1.60	0.27
	Class	16.90	1	16.90	2.84
	Sex x Class	0.90	1	0.90	0.15
	Error	214.20	36	5.95	—
	Total	233.60	39	—	—
Number of adjectives and adverbs	Sex	24.03	1	24.03	1.39
	Class	15.63	1	15.63	0.91
	Sex x Class	5.63	1	5.63	0.33
	Error	619.70	36	17.21	—
	Total	664.97	39	—	—

extent, to overcome the handicaps in language ability resulting from their underprivileged environments, apparently they do not acquire the same facility with difficult and unusual words which their more fortunate peers have attained. This difference is the more striking since the poor children in this study are on the average slightly brighter than the well-to-do children (Mean I.Q.'s 141 and 137).

The finding that boys are superior to girls is contrary to the findings of the majority of previous investigations, and may in part be explained by the fact that the boys were superior to the girls in intelligence, as indicated by differences in average I.Q.'s (143 and 135).

The major implication of this study rests in the obvious conclusion that schools must enrich the experiences of underprivileged children, whether they be dull or bright. The enrichment program should be concerned with verbal experiences and language skills of all kinds.

REFERENCES

1) Frank Auld, Jr., "Influence of Social Class on Personality Test Responses," *Psychological Bulletin*, 49: 318–332, 1952.

2) Helen H. Davidson, *Personality and Economic Background*. New York, King's Crown Press, 1943.

3) Allison Davis, *Social Class Influences Upon Learning*. Cambridge, Harvard University Press, 1948.

4) M. I. Dundson and J. A. F. Roberts, "The Relation of the Terman-Merrill Vocabulary Test to Mental Age in a Sample of English Children," *British Journal of Statistical Psychology*, 6: 61–70, 1953.

5) Erwin J. Lotsof, "Intelligence, Verbal Fluency, and the Rorschach Test," *Journal of Consulting Psychology*, 17: 21–24, 1953.

6) Kurt B. Mayer, *Class and Society*. Garden City, N. Y., Doubleday and Co., 1955.

7) Esther Milner, "Effects of Sex Role and Social Status on the Early Adolescent Personality," *Genetic Psychology Monographs*, 40: 231–325, 1949.

8) Esther Milner, "A Study of the Relationship between Reading Readiness in Grade One School Children and Patterns of Parent-Child Interaction," *Child Development*, 22: 95–112, 1951.

9) Dorothea McCarthy, "Some Possible Explanations of Sex Differences in Language Development and Disorders," *Journal of Psychology*, 35: 155–160, 1953.

10) Dorothea McCarthy, "Language Development in Children," in *Manual of Child Psychology*, edited by L. Carmichael, second edition. New York, Wiley and Sons, 1954.

11) Meyer Rabban, "Sex-role Identification in Young Children in Two Diverse Social Groups," *Genetic Psychology Monographs*, 42: 85–158, 1950.

12) Lewis Terman, "Psychological Sex Differences," in *Manual of Child Psychology*, edited by L. Carmichael, second edition. New York, Wiley and Sons, 1954.

13) Florene M. Young, "An Analysis of Certain Variables in a Developmental Study of Language," *Genetic Psychology Monographs*, 23: 1–141, 1941.

AN EXPERIMENTAL COMPARISON OF NONSENSE, WORD, AND SENTENCE MATERIALS IN EARLY TYPING TRAINING[1]

LEONARD J. WEST

Air Force Personnel and Training Research Center
Chanute Air Force Base, Illinois

Commercial typing textbooks as well as practice materials locally prepared by certain military training installations use a variety of materials in the early keyboard learning stages of typewriting instruction. These materials consist either of: (a) isolated words, (b) ordinary sentences, (c) "location drill" sequences made up of nonmeaningful series of letters struck by individual fingers—or of various combinations of these three types of materials.

Materials of the "location drill" or "nonsense drill" type (as they are characteristically called by typing teachers) were in use even before the pioneer experimental work of Book in 1908, republished subsequently together with a textbook for teachers (*1*, *2*). Their continued use on a very much reduced scale in today's commercial typewriting texts was somewhat supported by Book's description of typewriting skill as consisting of mastery over a hierarchy of stroking habits in which the typist first learns individual-letter habits before he can handle larger units of material in such a way that "the various movements required to write certain words and phrases become so closely linked together that the first movement in the series needs only to be purposely initiated and consciously controlled to set off all the movements in the series" (*1*, p. 173). Book's alternative suggestion that lower-order stroking habits might be perfected incidentally through the use of word materials or of connected discourse was reflected in a dozen attempts (during the period 1921–1948) to evaluate experimentally the merits of nonsense drill materials as against either

[1] This report is based on work done under ARDC Project No. 7714, Task No. 77260, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

word or sentence materials. A small but general tendency toward inferior performance on the part of those trained on nonsense materials in these experiments must be considered somewhat inconclusive largely because of failure to hold constant (a) the rate of introduction of new keys, (b) the frequency of appearance of each letter in the practice materials, and (c) the total amount of practice on each type of material.

The present experiment was designed to hold constant the three factors mentioned in comparing the merits of nonsense, word, and sentence materials as vehicles for the initial learning of key locations. In order to assess the possible influence of word length as a factor in difficulty, one set of isolated word materials was restricted almost entirely to those of three to five letters in length, while another set included words covering a wider range of length (two to seven letters) and averaged 0.5 letters per word more than those of the first set. With a view toward evaluating the four types of materials as preparation for ordinary sentence typing as well as for the typing of the random assortments of letters contained in the five-letter code groups typed by radio operators (e.g., *XIPKW LNUPM*), the final speed and accuracy of the subjects in this experiment were determined both for sentence and for code copy. While this experiment was not designed to be a direct test of any particular hypothesis underlying the nature of the practice materials and must be viewed as purely an empirical comparison of materials, certain factors which may account for the findings are discussed after the presentation of results.

METHOD

Materials. Samples of the four types of materials are shown in Figure 1 together with the practice schedule.

Several features of these materials should be mentioned. First, the nonsense sequences are *not* random assortments of letters, but fixed and rigid sequences specifically used for many years and designed to train each finger in turn to strike the keys assigned to it (e.g., only the left index finger is used for the sequence *frf*). Secondly, each line (or two) of the materials—and the ones shown are only samples—contains every letter of the alphabet. (The inclusion in the nonsense materials of certain characters and punctuation marks is demanded by the nature of the sequences. These are superfluous and were not included in the test materials.)

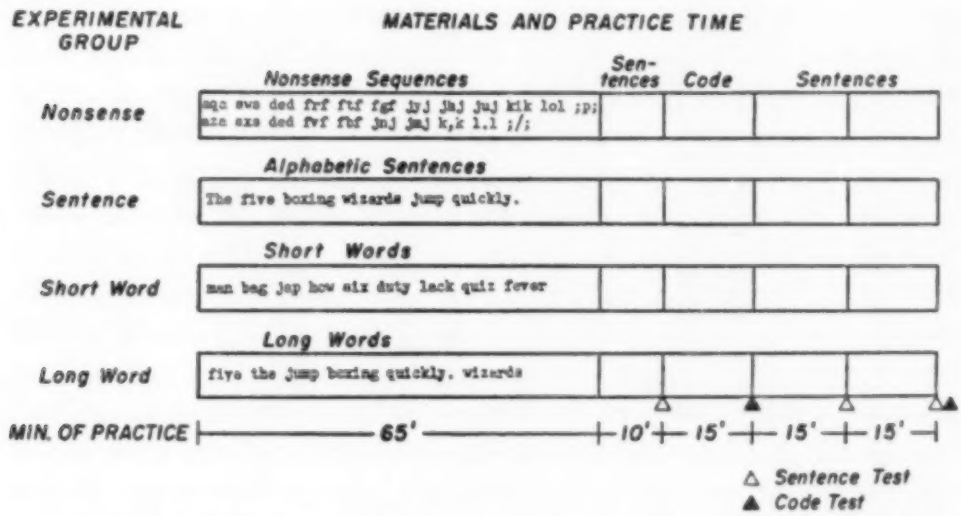

FIGURE 1. Minutes of typing practice on various materials by each of four experimental groups.

Third, it may be noted that the long word materials are simply the sentence materials in scrambled, nonsentence order. This was designed to furnish at least one comparison between materials in which letter sequences, in addition to other frequency variables, were held constant. Fourth, on the basis of the entire set of practice materials, those for the three word and sentence sets had identical letter frequencies. Discrepancies for the nonsense sequence typists were more than made up for by the anticipated finding that, because of the ease with which nonsense sequences may be typed, two and one-half times as much typing was done of the nonsense sequences as of any of the other three sets of materials.

Subjects. Three hundred forty-five airmen with no previous typing experience, either formal or informal, were used. Three-quarters of these men were awaiting entry into courses which include typing training. They were assigned to groups (one for each of the four types of materials) at random from within strata based on the course they were about to enter and on their Clerical Aptitude Index—the latter a measure used by the Air Force to predict success in clerical training and used as one basis of assignment to training. The nonsense, sentence, short word, and long word groups had, respectively, eighty-four, eighty, eighty-nine, and ninety-two men.

Procedure. The main procedural features of this experiment are shown in Figure 1. As shown, the first sixty-five minutes of a total

of one hundred and twenty minutes of practice were devoted to the differential materials. The last fifty-five minutes of practice used the same materials for all groups: ten minutes on a sentence designed to introduce capitalization to all learners; fifteen minutes on random, five-letter, equal-frequency code groups; and thirty minutes on sentences containing every letter of the alphabet. The same code and sentence materials were used by all groups in the second stage of practice in order to make the training more nearly correspond to standard procedures in which nonsense drill or word practice is envisioned purely as introductory practice to be followed by the use of materials gradually approaching those used on the job. Criterion measures consisted of speed and error scores on either one or two two-minute timings on new code or new alphabetic sentence materials.

Instructional sessions took place on four consecutive mornings for each of three weeks at the same time each day, in the same room, with the same equipment, and with the same instructor. A different group of men was used for each of the twelve sessions; there were, thus, three "subgroups" for each type of materials. Materials were assigned to groups at random within each four-day set of sessions, each session being devoted to a different one of the four treatments. Tape recordings of all instruction revealed virtually verbatim correspondence of instructor activities and comments throughout.

Each session lasted four hours: two hours of practice, sixteen minutes of testing, the remaining time being devoted to periodic rest breaks and to orientation and verbal and visual instructions. Substantial fatigue was noted during the latter stages of this relatively massed practice, terminal performance thereby being adversely affected.

Other relevant instructional aspects were as follows: (a) Except for the first ten minutes of practice, there was no repetitious practice on individual sequences, words, or even lines; units of six to ten different lines were typed before repetition was allowed. (b) During the initial stages of practice, subjects were urged to look at the keyboard as an aid to locating keys. Increasing emphasis on typing purely by touch was introduced after half an hour of practice. Various mixtures of sight and touch typing were observed to be in use by many learners through the first three-quarters of practice. Thereafter, and for the terminal measures, the mode of

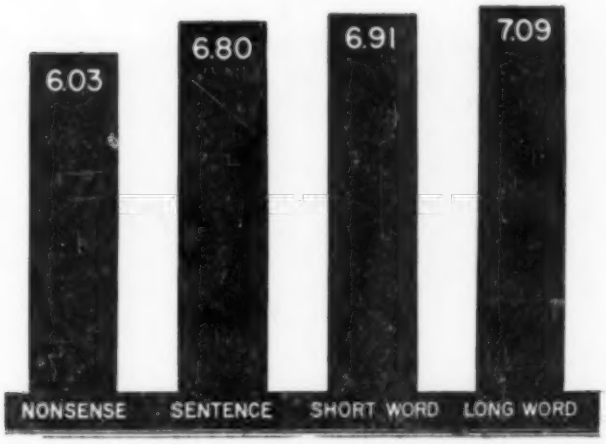

FIGURE 2. Mean words per minute of each of four groups on final, four-minute sentence test.

operation was observed to conform to standard touch typing procedures. (c) For the first line (or two) of practice (five to ten minutes) subjects were led through the materials step by step with explicit directions for each motion. All practice thereafter was at the learner's own rate.

RESULTS

Criterion measures consisted of total strokes (converted here to word-per-minute speeds) and number of errors on two two-minute final sentence tests and one two-minute final test on code materials.[2] In evaluating differences among and between groups, 0.05 was set as the minimum reliability level. Kruskal and Wallis' nonparametric analysis of variance by ranks (*3*, pp. 436–440) resulted in two chi-squares with probabilities indicating reliable differences among treatment ranks, namely, (a) 0.001 for sentence speed, and (b) 0.04 for code speed. That for sentence errors was

[2] Because the distributions of scores did not meet the assumptions concerning homogeneity of variance and normality of distribution desirable for analysis of variance, and because the effect of these discrepancies on analysis of variance was unknown, the data were subjected both to classical analysis of variance and to the nonparametric analysis of variance by ranks. The two types of analysis resulted in virtually identical probability levels. Over-all analyses were followed by t tests and by the nonparametric Mann-Whitney "U" test between various pairs of treatments—again producing virtually identical probability levels. In checking on consistency of treatment of the groups from session to session, it was found that there was no significant interaction between sessions and treatments.

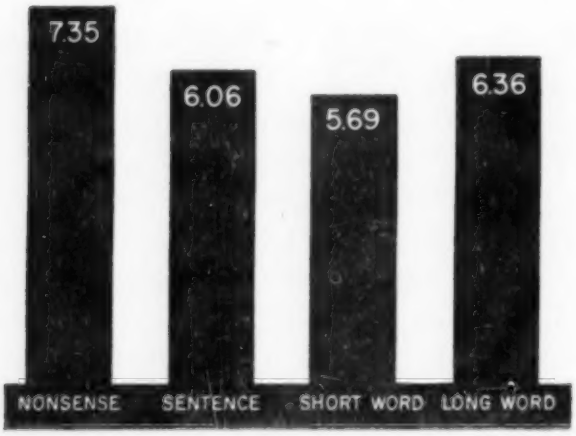

FIGURE 3. Mean number of errors of each of four groups on final, four-minute sentence test.

0.11, and there were no reliable differences for code errors. The mean performance scores for sentence speed, sentence errors, and code speed are shown in Figures 2, 3, and 4, respectively.

Although certain of the comparisons between pairs of treatments are neither orthogonal nor independent, those with probabilities exceeding 0.05 by Mann-Whitney "U" test (*3*, pp. 434–435) are shown in Table I.

Perhaps the most concise overview of the results may be seen in Table II, which displays the ranks of the four groups on each of the four criterion variables and the mean rank for each group.

It seems clear from these results that as measured at an early stage of training the initial use of the traditional, single-finger location drill sequences to teach key locations produces speeds which are inferior to those produced by learners trained on word

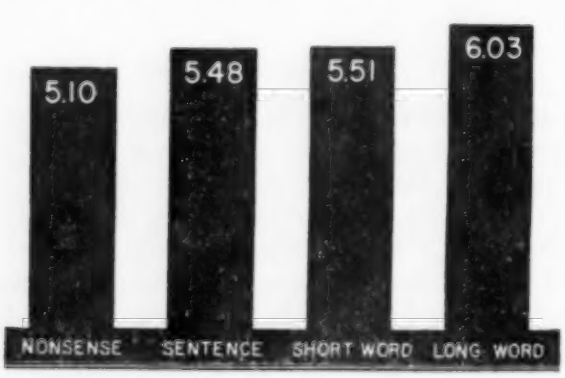

FIGURE 4. Mean words per minute of each of four groups on final, two-minute test on random, five-letter, equal-frequency code groups.

TABLE I.—PROBABILITY OF DIFFERENCE BETWEEN MEANS FOR VARIOUS PAIRS OF GROUPS BY MANN-WHITNEY "U" TEST

Experimental Pair*	Probability for		
	Sentence speed	Sentence errors	Code speed
Sentence-nonsense	0.01		
Short word-nonsense	0.001	0.02	
Long word-nonsense	0.0004		0.008

* The first-named group in each pair is the superior group. Only pairs for which differences are significant at the 0.05 level or better are shown.

TABLE II.—RANK OF MEAN SCORES FOR FOUR CRITERION VARIABLES FOR EACH OF FOUR EXPERIMENTAL GROUPS AND MEAN OF RANKS FOR EACH GROUP

Group	Rank on				Mean Rank
	Sentence		Code		
	Speed	Errors	Speed	Errors	
Nonsense	4	4	4	3	3.75
Sentence	3	2	3	2	2.50
Short word	2	1	2	1	1.50
Long word	1	3	1	4	2.25

or sentence materials. There is also an insufficiently reliable trend toward fewer errors on sentence materials by word and sentence typists. In general, both for sentence and code typing, the use of word or sentence materials appears to give learners an initial advantage over those trained on nonsense sequences.

DISCUSSION

Although this experiment was not designed to test directly possible reasons for the general superiority of word and sentence over nonsense materials, several speculations may be offered. It will be recalled that the stroking habits of typewriting were described as hierarchic, that nonsense materials by their very nature are designed to keep stroking on the lowest or letter-by-letter level. Combinations such as *sx*, *fv*, *jm*, etc., do not exist in the language and it does not seem likely that such combinations

would be handled in the grouped or patterned fashion which characterizes the use of higher-order stroking habits. In word and sentence materials, on the other hand, the learner practices the varied and characteristic sequences of common combinations of letters which the language contains. Such materials create opportunities, especially for the more able learner, to attempt, perhaps, a slight crowding of the more common letter sequences. The chief factor accounting for results is thus thought to be differences in the restrictiveness of the materials in allowing for individual differences in readiness to attempt higher-order stroking habits. A second consideration arises from the presence in nonsense materials of sequences which do not exist in the language. To whatever extent such sequences get stroked as sequences, they may interfere with the subsequent typing of meaningful materials and may have to be "unlearned." Third, the letter-combinations in location drill materials are very much fewer in number than those of dictionary materials. Thus, the word and sentence materials have the well-known advantage of greater variety of responses in the practice and thus greater transfer value than do the less varied responses of the traditional drill materials. Last, intrinsic interest factors are probably to the advantage of meaningful over nonsense materials.

SUMMARY

This experiment evaluated four types of early keyboard learning practice materials in typewriting: single-finger nonsense sequences, short words, long words, and sentences. Each of these four types of materials was used by a different group of learners for the first hour of practice. During the second and last hour of practice, all learners used the same code and alphabetic sentence materials. Statistical analysis of the speed and error scores of these learners on final tests using code and alphabetic sentence materials showed the word and sentence typists to be, in general, superior to those initially trained on traditional drill sequences. The chief factors which may account for the superiority of word and sentence over nonsense materials are thought to be: (a) the greater allowance for individual differences in ability to attempt stroking habits of a higher order, (b) the absence of possible interference from earlier practice on sequences which do not exist in the language, (c) the

greater variety of responses allowed by the practice, and (d) the higher levels of attention, concentration, and interest.

REFERENCES

1) W. F. Book, *Learning to Typewrite*. New York, Gregg Publishing Co., 1925.

2) W. F. Book, *The Psychology of Skill with Special Reference to Its Acquisition in Typewriting*. New York, Gregg Publishing Co., 1925.

3) H. M. Walker and J. Lev, *Statistical Inference*. New York, Henry Holt, 1953.

BEHAVIORAL CONTROL AND INTELLECTUAL ACHIEVEMENT OF SECONDARY SCHOOL BOYS[1]

NORMAN TALLENT[2]

Teachers College, Columbia University

Information on the relationships between the so-called non-cognitive personality variables on the one hand, and intellectual ability and vocational capacity on the other is often pertinent to problems of educational diagnosis and guidance. Factors of behavioral control, in particular, are thought to enter intellectual attainments, a low degree of behavioral control, or impulsiveness, being commonly thought to be incompatible with a high degree of intelligence. Webster's (*14*) defines "impulsive" as "impetuous," "quick," and "ill-considered." A psychological dictionary (*6*) apparently reflects the belief of most psychologists in stating that impulsiveness is characterized by a lack of ". . . foresight or prudence . . . in which the cognitive aspect is subordinated to the conative-affective aspects."

Some clinical evidence in support of this view has been accumulated (*10*), but relationships between behavioral control and intelligence have not been demonstrated in a school situation. Thurstone (*11*) reported correlation coefficients between a factor labeled "Impulsive" and some measures of intelligence. Correlations of scores on the Impulsive scale of the Thurstone Temperament Schedule with the total score and Q and L scores of the ACE, and with the total score and the subtest scores of the PMA, ranged from −0.073 to 0.152. All but the largest of these correlations, that between the Impulsive score and the PMA Word Fluency score, failed to reach the 0.05 level of significance.

---

[1] Based on a portion of the thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at Teachers College, Columbia University. Appreciation for assistance is extended to the writer's dissertation committee, Dr. Laurance F. Shaffer, Dr. Robert L. Thorndike, and Dr. Edward Joseph Shoben, Jr.

[2] Now at Veterans Administration Center, Kecoughtan, Virginia.

DEFINITION OF PROBLEM

The present study was designed to explore the relationship between behavioral control in school and performance on a wide range of intelligence tests. If it is agreed that behavioral control is itself an expression of intelligence, then a positive relationship should exist. Consideration was also given to the possibility that a greater relationship exists between behavioral control and intellectual achievement on some tests than on others. With this orientation, the following hypotheses were formulated:

(1) Subjects who are rated as showing a high degree of behavioral control have significantly higher mean intelligence test scores than subjects who are rated as impulsive.

(2) The above differences are significantly greater on tasks presumed to demand concentration, a high order of conceptualization, or evidence of long-term investment in intellectual matters, such as is seen in the accumulation of knowledge, than on tasks judged not to demand these qualities.

ASSESSMENT OF BEHAVIORAL CONTROL

A continuum of behavioral control was postulated with a high degree of control at one end, an average degree of control in a central position, and impulsiveness at the other. Nine supposed forms of undercontrolled behavior were then gleaned from an authoritative dictionary (*14*), from extensive psychoanalytic discussions of undercontrol (*5*, *9*), and from an examination of items on the Thurstone Impulsive Scale (*12*). Together with their opposite polarities, these forms of behavior may be listed as:

(1) Slowness of action vs. quickness of action.

(2) Slowness of thought vs. quickness of thought.

(3) Thought orientation vs. action orientation.

(4) Patience and persistence vs. impatience and lack of persistence.

(5) Carefulness and conscientiousness vs. carelessness and lack of conscientiousness.

(6) Conservative spirit vs. adventurous spirit.

(7) Socially positive behavior vs. social delinquency.

(8) Rigidity of procedure and thought vs. lability of procedure and thought.

(9) Stability of mood vs. lability of mood.

BEHAVIORAL CONTROL RATING SCHEDULE

A provisional Behavioral Control Rating Schedule (BCRS) was constructed, consisting of sixteen rating scales which were based on the above nine continua of behavioral control. Seven of the continua were represented by two scales each; two continua, speed of action and thought, were represented by one scale each in view of the consideration that intelligence might strongly influence speed. The first eight scales were intended to refer to approximately the same aspects of behavior as the last eight scales. Each scale was composed of four descriptive statements, having reference, insofar as possible, to concrete aspects of behavior. The statements of each scale extend from a high degree of behavioral control to undercontrol.

To use the schedule, one statement on each scale is checked. The number of the statement checked, 1, 2, 3 or 4 is taken as an index of control with reference to the aspect of behavior measured by the scale. The sum of the numbers checked on all the scales yields a figure which constitutes an index of general behavioral control.

STATISTICAL QUALITIES OF THE BCRS

To test the statistical qualities of the schedule, ratings were accomplished on eighty-six ninth-grade boys from the Elgin, Illinois, school system, as follows: (1) students and teachers in separate meetings with the experimenter were acquainted with the BCRS and asked to check on a mimeographed class roster the names of the students they felt they could rate on the schedule, (2) each student and each teacher completed ratings on five or fewer students from the names they had checked. In this manner, an average of 4.63 peer ratings and 4.07 teacher ratings were obtained on each student.

The corrected split-half reliability of the schedule, as rated by the students was 0.77. The teachers showed greater consistency in their ratings, a reliability coefficient of 0.90 being found.

The several ratings on each student were averaged and estimates of the between-rater reliabilities of the averaged multiple ratings of the students were computed by procedures based upon analysis of variance (*4*). The reliability of the averaged ratings was 0.86 for the student raters and 0.92 for the teacher raters. It thus appeared that averaged ratings of the BCRS could reliably be used

to separate students into different degrees of behavioral control as defined by this rating method.

A correlation between the averaged ratings of the students and the averaged ratings of the teachers was also computed. An r of 0.80 was obtained, indicating that these two groups of raters tended to evaluate the behavior of students similarly with the BCRS.

To determine the degree to which the various scales contribute to an estimate of behavioral control, an item analysis was carried out on a sample of sixty schedules completed by the teachers and sixty schedules completed by the students. The two scales which refer to speed of response were negatively correlated with the other postulated aspects of behavioral control, i.e., with the full BCRS, and were eliminated from the schedule. Otherwise, the various scales tend to contribute substantially to the full score. Scale F, as rated by the students, was anomalous in this respect, apparently because of some misleading wording which was retrospectively discovered and altered. The items of the revised BCRS are listed as Table I, the figures in parentheses preceding each scale referring to the eta correlation of that scale with the full score. The top figure is that of the student raters; the bottom figure is that of the teachers.

DESIGN AND CONDUCT OF THE EXPERIMENT

Classification of subjects into experimental groups. The subjects were one hundred and ninety-eight ninth-grade boys of the Elgin, Illinois, public school system. The degree of behavioral control of these subjects was appraised in terms of the averaged BCRS ratings by their teachers. An attempt was made to secure five ratings for each subject, but practical consideration reduced the number obtained to an average of 4.33. From the one hundred and ninety-eight rated, eighty were selected on the basis of the mean BCRS scores assigned them and placed in two groups. The forty subjects who received the highest averaged rating scores were designated the "Impulsive Group" (I Group) and the forty subjects who received the lowest averaged rating scores were labeled the "Controlled Group" (C Group). The mean rating scores for the I and C groups respectively were 41.73 and 25.05. It is noted that "impulsive" refers to the most poorly controlled subjects in a sample of normal school boys and the findings on

this group cannot be applied to persons having various degrees of pathological impulsiveness.

The intelligence test battery. The battery was selected for the purpose of testing the hypotheses of the study on a variety of intellectual tasks which represent current techniques for measuring intelligence. The battery included group tests and individual tests, verbal tests and performance tests, tests based on the factorial isolation of mental abilities, and tests which purport to measure intelligence without the intrusion of cultural influences. A number of orientations to intelligence influenced the design of the various tests. The tests were:

(1) Wechsler-Bellevue Intelligence Scale, Form I (*15*).

(2) Porteus Mazes, Vineland Edition (1933) (*8*).

(3) American Council on Education Psychological Examination for High School Students, 1948 Edition (ACE) (*1*).

(4) SRA Primary Mental Abilities for ages 11–17 (PMA) (*13*).

(5) A Culture-Free Test (*2*).

Classification of tests. All of the intelligence tests used in the study were examined by the experimenter in terms of the considerations which led to the formulation of Hypothesis 2. Tests which were judged to put impulsive subjects at a disadvantage, and hence show a high positive relationship between behavioral control and intelligence, were labeled "Relationship" or R tests. Tests which were judged not to put impulsive subjects at a disadvantage, and hence might show little or no relationship between intelligence and behavioral control, were labeled "No Relationship" or N tests.

All but seven tests were thus hypothesized to show a high relationship. Five of these were the performance tests of the Wechsler scale. The anticipated quickness in making decisions by the impulsive subjects, along with the ready opportunity afforded by the Wechsler performance tasks for the spontaneous correction of errors, were considered to be factors which would cause the scores of impulsive subjects to approximate those of other subjects. The Digit Span test was not expected to show relationship because of a clinical observation (*9*) that success on the test demands essentially a passive attitude. It was expected that impulsive subjects would also be enabled to score at approximately the level of their better controlled peers on the Word Fluency test. The underlying

TABLE I.—SCALES OF THE REVISED BEHAVIORAL CONTROL RATING SCHEDULE

A. — 1. Thinks very carefully before doing anything.
— 2. Gives at least some thought to a matter before doing (0.78) something about it.
(0.79) — 3. Tends to do something about a matter before he has a chance to give it much thought.
— 4. Acts before he has a chance to think at all about what he does.

B. — 1. Sticks to a task no matter how difficult.
(0.60) — 2. Gives up or quits only on very difficult tasks.
(0.77) — 3. Gives up or quits difficult tasks fairly quickly.
— 4. Gives up or quits a task as soon as he runs into any difficulty.

C. — 1. Does very careful, accurate work.
(0.63) — 2. Work is fairly careful and accurate.
(0.82) — 3. Tends to be careless and inaccurate.
— 4. Does very careless, "sloppy," inaccurate work.

D. — 1. Will never take a chance.
(0.79) — 2. Might take a chance.
(0.81) — 3. Tends to take chances.
— 4. Takes reckless chances.

E. — 1. Is exceptionally well-behaved—never gets into trouble.
(0.84) — 2. Is fairly well-behaved—might get into trouble.
(0.87) — 3. Gets into trouble quite often—not very well-behaved.
— 4. Is always getting himself into trouble—very badly behaved.

F. — 1. Can't be talked into changing his mind.
(−0.36) — 2. Will change his mind if given good reasons.
(0.63) — 3. Is quick to take or the ideas of others and change his mind.
— 4. Immediately takes on everyone else's ideas—doesn't have a mind of his own.

G. — 1. Does not seem to show his feelings (anger, sadness, happiness, etc.) at all.
(0.58) — 2. Does not show his feelings very easily.
(0.50) — 3. Shows his feelings fairly easily.
— 4. Shows his feelings very easily.

H. — 1. Spends his free time in serious conversation or study—definitely doesn't care to take part in sports or active games.

TABLE I—*Continued*

(0.44) — 2. Prefers serious conversation or study to taking part in sports or active games.
(0.79) — 3. Prefers taking part in sports or active games to serious conversation or study.
— 4. Spends his free time in taking part in sports or active games—doesn't care at all for serious conversation or study.

I. — 1. Always shows much patience.
(0.42) — 2. Generally shows a fair amount of patience.
(0.81) — 3. Tends to be rather impatient.
— 4. Is always very impatient.

J. — 1. Is very responsible—can be trusted to do anything.
(0.77) — 2. Is fairly responsible—can generally be trusted.
(0.82) — 3. Tends to be irresponsible—probably wouldn't care to trust him.
— 4. Is very irresponsible—can't be trusted to do anything.

K. — 1. Keeps away from action and excitement.
(0.49) — 2. Tends to keep away from action and excitement.
(0.78) — 3. Tends to be where there is action and excitement.
— 4. Is always in the middle of action and excitement—will find him where something is going on.

L. — 1. Always obeys the rules.
(0.91) — 2. Usually obeys the rules.
(0.86) — 3. Usually does not obey the rules.
— 4. Doesn't obey the rules at all—acts as if they don't apply to him.

M.— 1. Finds it very hard to change his ways of doing things—does things his own way or not at all.
(0.15) — 2. Can be talked into changing his ways of doing things.
(0.52) — 3. Readily agrees to another's way of doing things.
— 4. Would just as soon do something someone else's way as his own way.

N. — 1. Is a very serious person.
(0.71) — 2. Is a rather serious person.
(0.79) — 3. Tends to be carefree and gay.
— 4. Is very carefree and gay—really lets himself go and have a good time.

rationale was the expectation that impulsive subjects would make rapid, if not critical, associations, whereas it was thought that better controlled subjects would achieve their scores as a result of having superior verbal skills. The low positive correlation found by Thurstone (*11*) between Impulsive scores and Word Fluency scores further supported the decision to classify Word Fluency as an N test.

Testing program. The three group tests of intelligence (PMA, ACE, Cattell Culture-Free) were administered to one hundred and ninety-four students on whom ratings were available in separate testing sessions. The two individual tests of intelligence (Wechsler-Bellevue I, Porteus Mazes) were administered to all of the subjects comprising the C and I groups.

FINDINGS AND IMPLICATIONS

The data[3] indicate that there was a high relationship between intelligence scores and behavioral control ratings. There was a strong tendency for the subjects rated as showing a high degree of control to obtain the highest intelligence scores, with subjects rated as inferior on control achieving the lowest scores. Thus, the C Group means were higher than those of the I Group on twenty-five of twenty-eight subtests at or beyond the 0.01 level of confidence and on two of twenty-eight subtests at or beyond the 0.05 level.[4] Correlations between behavioral control scores and intelligence also show a pervasive relationship between these variables (Table II). A high score on the BCRS represents low control; therefore the negative correlations between control scores and intelligence scores indicate that the well-controlled subjects tended to obtain the highest intelligence scores and that the impulsive subjects tended to achieve the lowest scores. Hypothesis 1 is therefore clearly supported.

An obvious question arises in connection with these findings.

[3] To save printing costs, a table giving the mean scores and differences between means (together with t values for these differences) for the two groups on the tests of the intelligence battery has been deposited with the ADI Auxiliary Publications Project. Order Document No. 5099 from: Chief, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for photoprints, or $1.25 for 35 mm. microfilm.

[4] A group of boys rated average in behavioral control was also tested. The C Group means were significantly higher than those of the Average Group on most of the subtests and the Average Group tended to score higher than the I Group.

TABLE II.—CORRELATIONS BETWEEN SCALES OF THE BEHAVIORAL CONTROL RATING SCHEDULE AND GROUP TESTS OF INTELLIGENCE
(N = 194)

Test	BCRS Scale							
	Total	A	B	C	D	E	F	G
ACE	−0.45	−0.46	−0.51	−0.47	−0.34	−0.29	−0.33	−0.12
PMA	−0.37	−0.43	−0.52	−0.49	−0.27	−0.21	−0.33	−0.19
Culture-Free	−0.52	−0.50	−0.54	−0.58	−0.41	−0.45	−0.30	−0.19
	H	I	J	K	L	M	N	
ACE	−0.47	−0.24	−0.38	−0.24	−0.37	−0.26	−0.17	
PMA	−0.29	−0.15	−0.33	−0.10	−0.26	−0.28	−0.18	
Culture-Free	−0.42	−0.34	−0.45	−0.23	−0.47	−0.16	−0.36	

Could the teachers' ratings have been influenced by their knowledge of the students' scholarship, overestimating the behavioral control of the students who achieve academic success and underestimating the control of students who do not? The data of Table II can be interpreted to suggest that some errors were obtained from this source. The rating scales having the highest correlations with the group tests appear logically to concern forms of behavior which reflect affective personality components. Thus, the average correlations of Scales B, C, A and H with the group tests ranged from 0.39 to 0.52. These scales concern, respectively, ability to persevere at a task, carefulness and accuracy of work, tendency to think before acting, and relative preference for serious conversation or study to sports or active games. The average correlations of Scales G, K, M and N with the group tests ranged from 0.17 to 0.24. These scales concern, respectively, control over expression of feelings, tendency to be where there is action and excitement, readiness with which preferred modes of behavior are surrendered, and seriousness of mood. It is to be noted, however, that most of these low correlations are statistically significant. For the size of the sample employed, a correlation of 0.14 is significant at the 0.05 level and a correlation of 0.18 is significant at the 0.01 level. The correlation of 0.80 between teacher ratings of students and student ratings of their peers suggests further that the obtained

results are only partially based on teacher halo effect. The students showed good agreement with their teachers concerning the behavioral control of their fellow students even though it is unlikely that the students would be as cognizant of, or as concerned with, the scholarship of most of their peers as were the teachers.

The second hypothesis of this study was that the differences in mean scores between the C and I groups would be significantly greater on tasks judged to demand concentration, a high order of conceptualization, or evidence of long term investment in intellectual matters, such as is seen in the accumulation of knowledge (R tests), than on tasks judged not to demand these qualities (N tests). Such a finding would indicate that the R tests have a higher relationship to behavioral control than do the N tests.

This hypothesis was tested by the Mann-Whitney U test (7) which was used to estimate the probability that the t values for the differences between R test means and the t values for the differences between N test means were samples from a common population. The t values for the differences in means between the C and I groups were ranked from smallest to largest. The sum of ranks of the t values for the differences between the R test means was found to be proportionately greater than the sum of ranks of the t values for the differences between N test means. The obtained distribution of t values was found to be such as could be expected to occur by chance three times in a thousand if these t values were samples from a common population. Therefore, it was concluded that there were greater differences between the means of these two groups on R tests than on N tests.

To better define the patterns of intellectual functioning, the relative strengths and weaknesses among the two groups were analyzed. An additional advantage to such analysis is that it shows the relationship between behavioral control and intellectual achievement in the various areas measured by the battery independently of general intelligence levels, hence uncontaminated by possible rater bias which stems from knowledge of the subjects' school achievement. This operation was accomplished as follows: (a) subtest scores of the group tests were converted to standard scores (this operation was not necessary for the Wechsler scale because weighted Wechsler subtest scores are standard scores); (b) the deviations of each individual's subtest scores from the average of all of his subtest scores of each scale were computed;

TABLE III.—MEAN DEVIATIONS OF SUBJECTS' SUBTEST SCORES FROM THEIR AVERAGE SUBTEST SCORES AND DIFFERENCES BETWEEN THESE DEVIATIONS

Test	$\bar{D}_C$	$\bar{D}_I$	$\bar{D}_C - \bar{D}_I$
W-B Arithmetic	−0.05	−2.22	2.17**
W-B Similarities	0.77	−0.17	0.94*
W-B Picture Completion	0.62	1.44	−0.82*
W-B Digit Symbol	−1.15	0.29	−1.44**
PMA Space	−1.71	2.65	−4.36*
Cattell Series	2.94	−0.06	3.00*

* Significant at or beyond the 0.05 level.
** Significant at or beyond the 0.01 level.

(c) the deviations on each subtest for all subjects in each group were added algebraically, and the mean deviations for each group were computed; and, (d) the differences in mean deviations between groups were computed and tested for significance (t). Table III shows the mean deviations and the differences between deviations for all subtests which showed significant differences (0.01 or 0.05) when the mean deviations between groups were compared.

The number of tests on which a significant relationship was found between behavioral control and mean deviations was greater than might be expected on a chance basis if there were no real differences between groups in mental functioning. In considering the practical significance of these findings, it is noted that most of the tests are factorially complex when administered to school children (*3*). Hence, for the most part, the relationships between behavioral control and intellectual achievement demonstrated here are not relationships among neatly isolated psychological traits. Similarly, we may be quite sure that a number of factors contribute to achievement in any given school subject. The educational implications of the obtained relationships between behavioral control and intellectual achievement might therefore be examined to determine if any descriptive similarities may be found between performance on the tests in Table III and functioning in various school subjects or curriculum areas.

The Controlled Group shows superiority on the Wechsler Arithmetic test, a finding which is consistent with Rapaport's hypothesis that the ability to concentrate contributes to success on this test (*9*). Since the basic operations required on the Arithmetic test

enter the various quantitative skills, especially mathematics and science, the implications seem clear. The Controlled Group is also strong on the Similarities test and on the Series test, both tasks appearing to require considerable reasoning, including the use of abstract processes.

The Impulsive Group is relatively proficient on the Picture Completion test, a task which seems to call for an awareness of the environment, the ability and willingness to report evidences of such awareness, and very little intervening mental manipulation of data. These are minimal requisites for functioning in all school subjects. Of the more positive skills, the relative superiority of the Impulsive Group on the Space test indicates some ability to imagine the positions of objects in space. Findings on the Digit Symbol test suggest that, relative to intelligence, the Impulsive boys are superior on tasks of sensorimotor speed and coördination. There is reason to infer, therefore, that behaviorally well-controlled boys function at their best in areas demanding quantitative and abstract reasoning—areas in which more poorly controlled boys are relatively handicapped—and that these more impulsive school boys have at least some of the intellectual abilities which enter various forms of mechanical work and possibly also such tasks as engineering drawing and blueprint reading.

The design of the experiment makes it likely that the number of relations found here between behavioral control and intellectual achievement, when data are analyzed independently of general intelligence level, is minimal. It is possible that some real relationships were obscured by the complexity of factors entering performance on any given task, with competing tendencies giving rise to scores which fail to show significant differences when the C and I groups are compared. For example, the data analysis did not support the belief that poorly controlled boys are superior in Word Fluency. It is possible, however, that rapid, uncritical associations contributed to the scores of those low in control, this advantage being offset by the superior vocabulary of the well controlled.

The obtained relationships can not, of course, demonstrate causality or direction. It appears plausible that impulsive modes of behavior can detract from intellectual achievement, but it is also possible that, at least in some cases, behavioral undercontrol is based on lack of adequate school achievement—an observation

of many teachers. Where emotional factors are primary in detracting from intellectual achievement, counseling might be of maximum value in releasing potential. Where the student misbehaves due to frustration over nonachievement, course guidance might be most indicated.

SUMMARY

Relationships between the behavioral control and intellectual achievement of secondary school boys were investigated in view of the pertinence of such relationships to problems of educational diagnosis and guidance. Behavioral control was defined in terms of ratings by teachers on the Behavioral Control Rating Schedule, an instrument developed for this study. Intellectual achievement was determined by administering an extensive battery of intelligence tests. The major finding was that a pervasive relationship exists between behavioral control and intelligence, those boys rated as highest in control having the highest intelligence. The degree of relationship, however, was significantly higher on some tests than on others. Certain relative strengths and weaknesses were also found to be associated with different levels of behavioral control. Some of the possible implications of these findings for student guidance were discussed.

REFERENCES

1) *American Council on Education Psychological Examination for High School Students*, 1948 edition. Princeton, N. J., Educational Testing Service, 1948.

2) R. B. Cattell, *A Culture-Free Test*. New York, The Psychological Corp., 1944.

3) P. C. Davis, "A Factor Analysis of the Wechsler-Bellevue Intelligence Scale, Form I, in a Matrix with Reference Variables," *American Psychologist*, (Abstract). 7: 296–297, 1952.

4) R. L. Ebel, "Estimation of the Reliability of Ratings," *Psychometrika*, 16: 407–424, 1951.

5) O. Fenichel, *The Psychoanalytic Theory of Neurosis*. New York, Norton, 1945.

6) P. L. Harriman, *The New Dictionary of Psychology*. New York, Philosophical Library, 1947.

7) H. B. Mann and D. R. Whitney, "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics*, 18: 50–60, 1947.

8) S. D. Porteus, *Porteus Mazes, Vineland Edition (1933)*. New York, The Psychological Corp., 1933.

9) D. Rapaport, M. Gill and R. Schafer, *Diagnostic Psychological Testing*, 2 volumes. Chicago, Year Book Publishers, 1945.

10) R. Schafer, *The Clinical Application of Psychological Tests*. New York, International Universities Press, 1948.

11) L. L. Thurstone, *Examiner Manual for the Thurstone Temperament Schedule*. Chicago, Science Research Associates, 1950.

12) L. L. Thurstone, *Thurstone Temperament Schedule*. Chicago, Science Research Associates, 1949.

13) L. L. Thurstone and Thelma G. Thurstone, *SRA Primary Mental Abilities*. Chicago, Science Research Associates, 1947.

14) *Webster's New International Dictionary of the English Language*, second edition, unabridged. Springfield, Mass., G. & C. Merriam, 1934.

15) D. Wechsler, *Wechsler-Bellevue Intelligence Scale* (Form I). New York, The Psychological Corp., 1947.

BOOK REVIEW

V. J. McGILL. *Emotions and Reason*. Springfield, Ill.: Charles C Thomas. 1954. pp. 122.

When a man loses his head, he acts stupidly, his conduct is described as emotional. When a man thinks clearly and behaves in a deliberate and rational fashion, it is assumed that emotion is absent or suppressed or under control. This the author does not agree with—hence the book on *Emotions and Reason* to prove the need for applying what he considers to be a more modern approach —elimination of the sharp division between emotion and reason.

His approach, the author believes, will eliminate the split between theory and practice. In this respect, he quotes John Dewey as saying that this split "is probably one of the chief sources of the maladjustments and unendurable strains from which the world is suffering." The author ends the book with an expression of hope that this short book of his will help in the general trend that he sees developing of the functional theory of emotions that will help the opponents of dualism.

Generally speaking, the book is expressed in rather abstract language throughout. In his attempt to summarize the present literature on emotions or reason, the author depends largely on philosophical background for interpreting and discourse. The argument is presented under three chapters: Needs, Emotions and Ethics, and the Emotions. For his background and for the words he uses in discussing the problems, the author gives credit to and shows familiarity with the writings of his former teachers, Edmund Husserl, G. E. Moore and Ralph Barton Perry. Psychologists he particularly likes to quote are: E. C. Tolman and D. O. Hebb.

H. MELTZER

Psychological Service Center, St. Louis, Mo.

(Concluded from inside front cover)

these charges, as well as information regarding the services of the American Documentation Institute in making available tables, etc., that for economic reasons may be omitted from a printed article, by writing to the Editors or the Publishers.

Double-spacing.—Manuscripts should be typed, written on one side of the paper only, and double-spaced throughout including quotations, footnotes, and bibliographical references.

Footnotes.—Footnotes are to be numbered consecutively beginning with '1', and should be on a separate sheet at the end of manuscript. (Footnotes to tables carry the *, †, and ‡.)

Titles.—Titles of articles should be brief, preferably three to eight words, with an extreme maximum of twelve words.

Type style.—Manuscripts are not to be marked for type style—this is done in the editorial office.

Books and other materials for review, and business communications should be addressed to THE JOURNAL OF EDUCATIONAL PSYCHOLOGY, Warwick & York, Publishers, 10 E. Centre St., Baltimore 2, Md.

Subscribers should notify the Publishers of change of address at least four weeks in advance of publication of the issue with which change is to take effect; both the old and new address should be given.

The Publisher desires every subscriber to get all issues to which he is entitled. Each journal is securely enclosed in a sturdy wrapper on which the subscriber's name and address have been printed, and is delivered directly to the Post Office, postage prepaid. Second-class matter is handled less by postal employees than other mail; moreover, if the Post Office is unable to make delivery, a notice to this effect is sent the Publisher and the magazine returned. Consequently, it is doubtful if one journal in many thousands is actually lost in transit.

But after an issue has been delivered to the proper address many things may happen to it—it may be diverted, or misplaced, or borrowed and not returned. For this neither Post Office nor Publisher is responsible. However, a subscriber who does not find a given issue in its assigned place may innocently make a claim of non-receipt. No claim for non-receipt of an issue can be honored unless made within four weeks after arrival of the next succeeding number. In order that a claim may arrive within the time limit it should be addressed to the Publisher—not to an agency.

WARWICK AND YORK *Publishers* BALTIMORE 2, MD.

E.C.

The Journal of Educational Psychology